Technical Note: The 45 and Up Study COVID Data Hub

Version: November 2022

> This technical note should be read in conjunction with the 45 and Up Study COVID Data Hub Data Dictionary, the 45 and Up Study Wave 3 Data Dictionary and the 45 and Up Study COVID Insights protocol paper [1].

Overview

The 45 and Up Study COVID Data Hub contains survey data for 61,941 45 and Up Study participants who answered questions about the impact of the COVID-19 pandemic on their lives. The dataset brings together responses from six surveys: the COVID Supplement from the 45 and Up Study Wave 3 follow up questionnaire in 2020, and the five surveys that make up the 45 and Up Study COVID Insights survey series. The COVID Data Hub includes 167,486 completed surveys. Geocoded address data for 32,064 of the COVID Insights survey participants is available separately.

Survey topics included health care utilisation, mental health and vaccination. Pragmatic modifications were made to some questions over the course of the surveys in order to reflect changing contexts and emerging needs.

This technical note includes information about recruitment, survey timing, demographic data, derived variables in the dataset and geocoded address data.

Recruitment timeline and survey tranches

The COVID Data Hub includes responses from the 45,078 people who completed the COVID Supplement as part of the 45 and Up Study Wave 3 follow up questionnaire in 2020 and the 32,115 people who participated in the 45 and Up Study COVID Insights project. 15,252 people participated in both (see Figure 1). 62% (n=20,027) of the COVID Insights participants completed all the surveys they were eligible for.

COVID Insights was an online survey only. The COVID Supplement as part of the 45 and Up Study Wave 3 follow up questionnaire in 2020 included postal and online options for responding.

Participants were recruited to the COVID Insights project at two different times [1]. Invitations were sent to the first 60,000 people (n=40,000 by email and n=20,000 by post) in November 2020. 16,863 people responded to this invitation and completed Survey 1 in November or December 2020.

Invitations to Survey 2 were split into two tranches. The first tranche (2a) consisted of the group that were recruited in November and December 2020. The second tranche (2b) consisted of invitations to

people who had expressed interest in participating in COVID-19 research when they completed their 45 and Up Study Wave 3 follow up questionnaire in 2020 (n=28,840). 15,252 people responded to this invitation and completed their first COVID Insights survey in March or April 2021.

Invitations to Survey 3 and 4 were each split into three tranches (a, b and c), in order to capture changing behaviours and sentiment over the period that each survey was in the field. Invitations to Surveys 1 and 5 were each sent in a single tranche. Table 2 outlines tranches, survey timing and response rates.

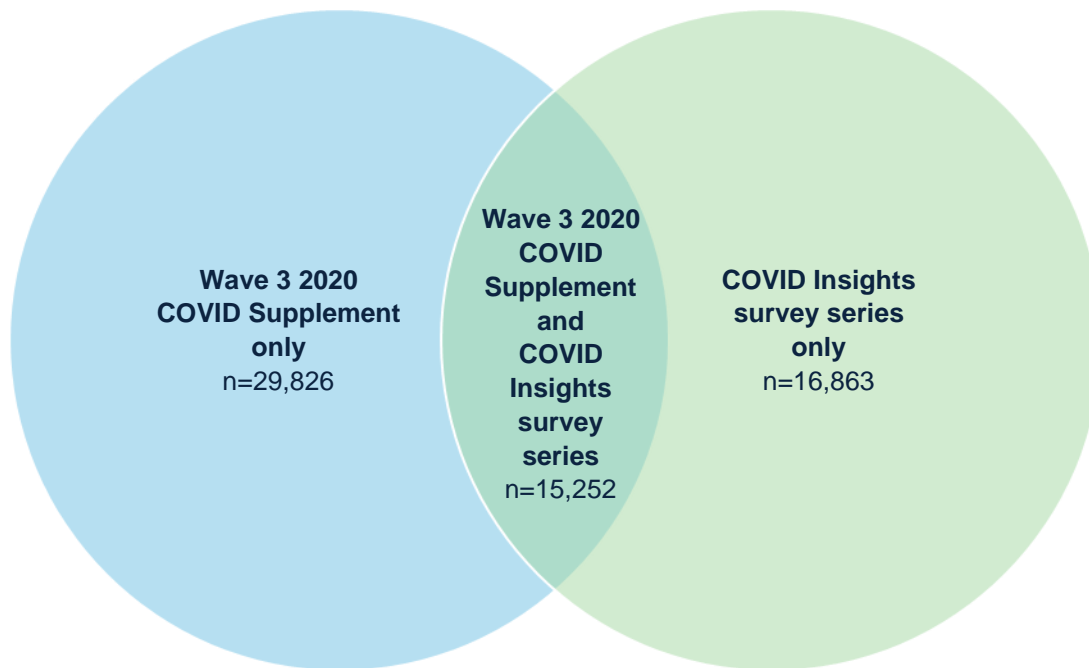**Figure 1**: Participant composition in the COVID Data Hub (n = 61,941).

Wave 3 2020 COVID Supplement only
n=29,826

Wave 3 2020 COVID Supplement and COVID Insights survey series
n=15,252

COVID Insights survey series only
n=16,863

Table 2: Dates in field and response to each survey in the COVID Data Hub

Survey	Tranche	Dates in field	Invited	Responded	Response rate
Wave 3 2020 follow up survey	-	31/07/2020 – 8/02/2021	85,299	45,078	52.8%
Survey 1*	-	02/11/2020 – 22/12/2020	60,000	16,863	28.1%
Survey 2	2a	18/02/2021 – 16/03/2021	16,839	13,724	81.5%
	2b*	10/03/2021 – 12/04/2021	28,840	15,252	52.9%
Survey 3	3a	10/06/2021 – 06/07/2021	14,000	11,027	78.8%
	3b	07/07/2021 – 03/08/2021	9,155	8,182	89.4%
	3c	04/08/2021 – 01/09/2021	8,863	7,806	88.1%
Survey 4	4a	08/09/2021 – 05/10/2021	13,750	10,781	78.4%
	4b	06/10/2021 – 02/11/2021	9,069	7,424	81.9%
	4c	03/11/2021 – 30/11/2021	8,742	6,512	74.5%
Survey 5	-	01/03/2022 – 01/04/2022	31,299	24,837	79.4%

* Denotes recruitment to the COVID Insights series at this survey.

Demographic data

Eleven demographic variables are included in the COVID Data Hub. Age was calculated at the time each participant completed each survey. The other 10 demographic variables (sex, CALD, language other than English spoken at home, disability status, number of chronic conditions, carer status, housing type, number of people in household, ARIA and SEIFA) were derived once all data collection was complete. Therefore, these ten variables are static within the COVID Data Hub for each person.

Derived variables

The derived variables included in the COVID Data Hub are listed below, and include three validated measures relating to mental health: the De Jong Gierveld Loneliness Scale (6-item short scale) [2], the Duke Social Support Index (abbreviated 10-item scale with two sub-scales) [3], and the Kessler Psychological Distress Scale (6-item short scale) [4].

Age

Age is given for each participant at each survey. Age was derived from the participant's date of birth and the date the survey was completed.

This derived variable is called 'age' in the dataset.

Cultural and linguistic diversity (CALD)

A participant's culturally and linguistically diverse (CALD) status was derived from their responses to two questions in the 45 and Up Study Baseline questionnaire (2005-2009). At Baseline, participants were asked which country they were born in, and whether they speak a language other than English at home.

If a participant was born in a country other than Australia, New Zealand, UK, England, Northern Ireland, Scotland, Wales, Ireland, Canada, USA or South Africa, then they were counted as CALD. Additionally, anyone who indicated that they speak a language other than English at home was counted as CALD.

This derived variable is called 'd_CALD' in the dataset.

Disability status

A standard question asked of 45 and Up Study participants is whether they regularly need help due to illness or disability. This question was asked at the Baseline questionnaire (2005-2009), the Social, Economic and Environmental Factors (SEEF) questionnaire (2010), Wave 2 (2012-2015) and Wave 3 (2018-2020). A participant's most recent response to this question was used to determine their disability status.

This derived variable is called 'd_disability' in the dataset.

Chronic conditions

A standard question asked of 45 and Up Study participants is whether a doctor has ever told them that they have a range of health conditions. This question was asked at the Baseline questionnaire (2005-2009), the Social, Economic and Environmental Factors (SEEF) questionnaire (2010), Wave 2 (2012-2015) and Wave 3 (2018-2020). A variation on this question was asked at COVID Insights surveys 3 and 5. If a participant responded 'yes' at any survey to any of the below sub-conditions, they were considered to have the overarching chronic condition (see Table 3). Each overarching category was counted as one chronic condition, regardless of how many of the sub-conditions the participant had been told they have. For example, if a participant said they had been told that they

had both melanoma and bowel cancer, this was counted as one overarching chronic condition – cancer.

This derived variable is called 'd_chronicconditions' in the dataset.

Details of chronic conditions and the age they were diagnosed are available by linking to other 45 and Up data collections.

Table 3: Health conditions that make up chronic conditions

Overarching condition	Sub-conditions (as asked in questionnaires)
Cancer	Melanoma, breast cancer, prostate cancer, bowel cancer, blood cancer, other cancer. (Note that non-melanoma skin cancer is not counted)
Cardiovascular disease	Heart or cardiovascular disease, angina, heart attack, heart failure (e.g. cardiac failure, weak heart, enlarged heart), rhythm disorder, valve disorder, high blood pressure (hypertension), peripheral vascular disease, atrial fibrillation
Stroke	Stroke
Diabetes	Diabetes, diabetes type 1, diabetes type 2 or unsure
Asthma	Asthma
Arthritis	Osteoarthritis
Parkinson's disease	Parkinson's disease
Chronic kidney disease	Chronic kidney disease

Carer status

A standard question asked of 45 and Up Study participants is whether they regularly care for a sick or disabled family member or friend. This question was asked at the Baseline questionnaire (2005-2009), the Social, Economic and Environmental Factors (SEEF) questionnaire (2010), Wave 2 (2012-2015), Wave 3 (2018-2020) and COVID Insights survey 3 (2021). Each participant's most recent response to this question was used to determine their carer status.

This derived variable is called 'd_carer' in the dataset.

Housing

A standard question asked of 45 and Up Study participants is what type of housing they live in. This question was asked at the Baseline questionnaire (2005-2009), the Social, Economic and Environmental Factors (SEEF) questionnaire (2010), Wave 2 (2012-2015), Wave 3 (2018-2020) and

COVID Insights survey 5 (2022). Each participant's most recent response to this question was used to determine their housing type.

This derived variable is called 'd_housing' in the dataset.

Living arrangements

A standard question asked of 45 and Up Study participants is how many people live in their household. This question was asked at the Social, Economic and Environmental Factors (SEEF) questionnaire (2010), Wave 2 (2012-2015), Wave 3 (2018-2020) and the COVID Insights survey 4 (2022). Each participant's most recent response to this question was used to determine their living arrangements.

This derived variable is called 'd_livingarrangement' in the dataset.

Remoteness (ARIA)

A participant's remoteness was determined based on their address at the time of COVID Insights survey 1 or the Wave 3 follow up survey in 2020. Each participant's most recently geocoded address was used to determine whether they live in a city, regional or rural area.

See Geocoded address data for more details.

This derived variable is called 'd_RA_NAME_2011' in the dataset.

Socio-economic disadvantage (SEIFA)

A participant's decile of socio-economic disadvantage was determined based on their address at the time of COVID Insights survey 1 or the Wave 3 follow up survey in 2020. Each participant's most recently geocoded address was used to determine this decile.

See Geocoded address data for more details.

This derived variable is called 'd_Decile_RSE_Disadvantage' in the dataset.

De Jong Gierveld Loneliness Scale

The six items that comprise the shortened version of the De Jong Gierveld Loneliness Scale [2] were asked at the Wave 3 2020 COVID Supplement and surveys 1, 3 and 5 of the COVID Insights series, in order to capture changes in loneliness over the course of the pandemic. An overall score for the 6-item scale was calculated, along with scores for the 3-item emotional and social loneliness sub-scales. Scale scores were only calculated when there were no missing values. The SAS code used to calculate the scores is as follows:

```
if survey in ('Wave3','1','3','5') then do;
     empty = .;
     if feelempty in ('1', '3') then empty = 1;
     else if feelempty = '2' then empty = 0;
     missppl = .;
     if feelmisspeople in ('1', '3') then missppl = 1;
     else if feelmisspeople = '2' then missppl = 0;
     reject = .;
```

```
            if feelrejected in ('1', '3') then reject = 1;
            else if feelrejected = '2' then reject = 0;
            rely = .;
            if feelplentypeopletorelyon in ('2', '3') then rely = 1;
            else if feelplentypeopletorelyon = '1' then rely = 0;
            trust = .;
            if feelpeopletrustcompletely in ('2', '3') then trust = 1;
            else if feelpeopletrustcompletely = '1' then trust = 0;
            close = .;
            if feelpeoplefeelcloseto in ('2', '3') then close = 1;
            else if feelpeoplefeelcloseto = '1' then close = 0;
end;

* De Jong Gierveld loneliness score;
array dejong {6} empty missppl reject rely trust close;
dejongnmiss = nmiss(of dejong(*));
if dejongnmiss = 0 then DeJongscore = sum(of dejong(*));
else DeJongscore = .;
* De Jong Gierveld emotional loneliness score;
array dejongemot {3} empty missppl reject;
dejongemotnmiss = nmiss(of dejongemot(*));
if dejongemotnmiss = 0 then DeJongemotscore = sum(of dejongemot(*));
else DeJongemotscore = .;
* De Jong Gierveld social loneliness score;
array dejongsoc {3} rely trust close;
dejongsocnmiss = nmiss(of dejongsoc(*));
if dejongsocnmiss = 0 then DeJongsocscore = sum(of dejongsoc(*));
else DeJongsocscore = .;
```

Duke Social Support Index

The four items that comprise the social interaction sub-scale of the Duke Social Support Index (DSSI)
[3] were included in the Wave 3 2020 follow up survey. All 10 items of the abbreviated DSSI were
included at survey 1 of the COVID Insights series. Response options differed between the Wave 3
follow up survey and the COVID Insights survey for the social interaction sub-scale, so different
variables were used for each. The Wave 3 follow up survey allowed participants to enter a numeric
response, whereas they had to choose a response option in COVID Insights survey. Wave 3
respondents have a score for the 4-item social interaction sub-scale, and COVID Insights survey 1
respondents have scores for the 10-item DSSI, the 4-item social interaction sub-scale and the 6-item
satisfaction with support sub-scale. Scale scores were only calculated when there were no missing
values. The SAS code used to calculate the scores is as follows:

```
* Wave3 respondents;
if survey='Wave3' then do;
        depend = .;
        if peopledependnum = 0 then depend = 1;
        else if peopledependnum in (1,2) then depend = 2;
        else if peopledependnum > 2 then depend = 3;
        visit = .;
        if socialvisittimes = 0 then visit = 1;
        else if socialvisittimes in (1,2) then visit = 2;
        else if socialvisittimes >= 3 then visit = 3;
        phone = .;
        if socialtelephonetimes in (0,1) then phone = 1;
        else if socialtelephonetimes in (2,3,4,5) then phone = 2;
        else if socialtelephonetimes >= 6 then phone = 3;
        group = .;
        if socialgrouptimes in (0,1) then group = 1;
        else if socialgrouptimes in (2,3,4,5) then group = 2;
        else if socialgrouptimes >= 6 then group = 3;
end;
```

```
* Survey 1 respondents;
if survey='1' then do;
        depend = .;
        depend = input(peopledependnum2, 8.);
        visit = .;
        if socialvisittimes2 = '1' then visit = 1;
        else if socialvisittimes2 in ('2','3') then visit = 2;
        else if socialvisittimes2 in ('4','5','6','7','8') then visit = 3;
        phone = .;
        if socialtelephonetimes2 in ('1','2') then phone = 1;
        else if socialtelephonetimes2 in ('3','4','5','6') then phone = 2;
        else if socialtelephonetimes2 in ('7','8') then phone = 3;
        group = .;
        if socialgrouptimes2 in ('1','2') then group = 1;
        else if socialgrouptimes2 in ('3','4','5','6') then group = 2;
        else if socialgrouptimes2 in ('7','8') then group = 3;

        understand = input(famfriendunderstand, 8.);
        useful = input(famfrienduseful, 8.);
        goingon = input(famfriendgoingon, 8.);
        listen = input(famfriendlisten, 8.);
        role = input(famfriendrole, 8.);
        problems = input(famfriendproblems, 8.);
end;

array dssiinteract{4} depend visit phone group;
array dssisupport{6} understand useful goingon listen role problems;
array dssiall{10} depend visit phone group understand useful goingon listen role
problems;

* Calculate social interaction sub-scale;
if survey in ('Wave3', '1') then do;
        if nmiss(of dssiinteract{*}) = 0 then DSSIsoci = sum(of dssiinteract{*});
        else DSSIsoci = .;
end;
* Calculate satisfaction with support sub-scale;
if survey = '1' then do;
        if nmiss(of dssisupport{*}) = 0 then DSSIsupt = sum(of dssisupport{*});
        else DSSIsupt = .;
* Calculate 10-item DSSI;
        if survey= '1' then DSSI = sum(of dssiall{*});
end;
```

Kessler Psychological Distress Scale

The six items that comprise the shortened version of the Kessler Psychological Distress Scale (K6) [4] were asked at the Wave 3 2020 follow up survey and surveys 1, 3 and 5 of the COVID Insights series, in order to capture changes in mental health over the course of the pandemic. A scale score was only calculated when there were no missing values. The SAS code used to calculate the scores is as follows:

```
array K6 (6) feelnervous feelhopeless feelrestless feeleffort feelsosad
feelworthless;
K6nmiss = nmiss(of K6(*)) ;
if K6nmiss = 0 then K6score = sum(of K6(*)) ;
else if K6nmiss >0 then K6score = .;
```

Geocoded address data

Geocoded address details can be supplied for 32,064 of the 32,115 COVID Insights participants. Only NSW addresses were geocoded. Addresses were geocoded using Callpoint Spatial. Participants who joined the COVID Insights project at survey 1 in November to December 2020 were geocoded based on their address at that time. Participants who completed the COVID Supplement from the 45 and Up Study Wave 3 follow up questionnaire in 2020 and then went on to join the COVID Insights project at survey 2 in March or April 2021 were not geocoded again when they joined COVID Insights. The addresses geocoded at Wave 3 2020 were used for these people. Not all addresses could be mapped by the geocoder (PO boxes, property names etc), so work was undertaken in-house to code these based on postcode so there was complete geocode coverage for the COVID Insights cohort.

References

1. Dawson G, Bleicher K, Baynes S, D'Este C, Steinberg J, Weber M, et al. 45 and Up COVID Insights: A dynamic and collaborative approach to evidence-making during the COVID-19 pandemic. Public Health Res Pract. September 2022; Online early.
2. De Jong Gierveld J, Van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing. 2010;7(2):121-30.
3. Koenig HG, Westlund RE, George LK, Hughes DC, Blazer DG, Hybels C. Abbreviating the Duke Social Support Index for use in chronically ill elderly individuals. Psychosomatics. 1993; 34(1):61-69.
4. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med. 2002;32(6):959-76.

Further information can be found in the COVID Data Hub Data Dictionary, or by contacting the 45 and Up Study Data Team by email on 45andUp.data@saxinstitute.org.au.